Geographical and Temporal Structures of *Legionella pneumophila* Sequence Types in Comunitat Valenciana (Spain), 1998 to 2013

Leonor Sánchez-Busó,[a,b] Mireia Coscollà,[a,c,d] Ferran Palero,[a,e] María Luisa Camaró,[f] Ana Gimeno,[f] Pilar Moreno,[g] Isabel Escribano,[h] María Mar López Perezagua,[i] Javier Colomina,[j] Herme Vanaclocha,[a,k] Fernando González-Candelas[a,b]

Joint Research Unit Infección y Salud, FISABIO-Universidad de Valencia, Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Valencia, Spain[a]; CIBER en Epidemiología y Salud Pública, Valencia, Spain[b]; Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland[c]; University of Basel, Basel, Switzerland[d]; INRA, University of Nice Sophia Antipolis, CNRS, UMR 1355-7254 Institut Sophia Agrobiotech, Sophia Antipolis, France[e]; Public Health Laboratory, Valencia, Spain[f]; Centro de Salud Pública de Alcoy, Alcoy, Spain[g]; Laboratorio de Microbiología, Hospital Virgen de los Lírios, Alcoy, Spain[h]; Laboratorio de Microbiología, Hospital Marina Baixa, La Vila Joiosa, Spain[i]; Laboratorio de Microbiología, Hospital Alzira, Alzira, Spain[j]; DGSP, Subdirección de Epidemiología y Vigilancia de la Salud, Valencia, Spain[k]

Legionella pneumophila is an accidental human pathogen associated with aerosol formation in water-related sources. High recombination rates make *Legionella* populations genetically diverse, and nearly 2,000 different sequence types (STs) have been described to date for this environmental pathogen. The spatial distribution of STs is extremely heterogeneous, with some variants being present worldwide and others being detected at only a local scale. Similarly, some STs have been associated with disease outbreaks, such as ST578 or ST23. Spain is among the European countries with the highest incidences of reported legionellosis cases, and specifically, Comunitat Valenciana (CV) is the second most affected area in the country. In this work, we aimed at studying the overall diversity of *Legionella pneumophila* populations found in the period from 1998 to 2013 in 79 localities encompassing 23 regions within CV. To do so, we performed sequence-based typing (SBT) on 1,088 *L. pneumophila* strains detected in the area from both environmental and clinical sources. A comparison with the genetic structuring detected in a global data set that included 20 European and 7 non-European countries was performed. Our results reveal a level of diversity in CV that can be considered representative of the diversity found in other countries worldwide.

The main reservoirs of *Legionella pneumophila* in natural habitats are water-related environments, where it is known to form biofilms in surface interphases (1). *L. pneumophila* can multiply actively within these complex structures (2), disperse through the water flow, and colonize different urban water distribution systems and risk facilities, such as cooling towers (3). Aerosols loaded with a sufficient amount of this bacterium contribute to its dispersal (4), and when inhaled by humans, it can produce an opportunistic pneumonia known as Legionnaires' disease (LD) (5). No person-to-person transmission has been reported for *Legionella*, so it is commonly considered a strictly environmental pathogen (4).

Different molecular markers have been used to characterize *Legionella* populations in previous genetic studies (6), but sequence-based typing (SBT) has become the current gold-standard typing tool (7, 8). Apart from its utility in epidemiological investigations of LD outbreaks (9–11), SBT provides nucleotide sequence data that can be used for further analyses of genetic variability and population structure (12, 13). A recent study has shown that even though undetected genomic variability within sequence types (ST) could mislead the identification of *Legionella* reservoirs during outbreak investigations, genetic distances determined by using SBT data correlate significantly with genome-wide estimates (14).

Characterizing the distribution of different STs across space and time is essential for a better understanding of dispersal patterns of environment-associated bacteria. This is particularly true for *L. pneumophila*, which is generally considered an accidental pathogen (15). Our team has been analyzing most of the strains detected during outbreak investigations and surveillance programs in the Comunitat Valenciana (CV) region for >15 years.

During the period from 1999 to 2011, CV and Cataluña represented over three-quarters of the total legionellosis outbreaks reported in Spain, with 124 and 331 registered episodes, respectively (16). Despite being the second most important Spanish region in the number of reported LD cases, *Legionella* populations in the CV region are still poorly characterized. An initial report using only three loci from the SBT scheme to genotype a few isolates from the Alicante province (within the CV region) already revealed a large variability within this area (12), comparable to that found in Europe.

In this work, we present an extensive survey of the genetic variability and population structure of *L. pneumophila* strains, including >1,000 isolates sampled throughout the whole CV region during the 1998-2013 period. In addition, we compared the genetic diversity detected in CV with data in the *Legionella*

pneumophila sequence-based typing database (http://www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php), which includes strains from 20 European and 7 non-European countries. Our objective is to provide a more thorough view of the genetic diversity of this opportunistic pathogen at different geographic scales, which can be used for epidemiological analyses in the area in subsequent years.

MATERIALS AND METHODS

Strain collection and processing. A total of 1,088 clinical and environmental samples were collected across the CV region during the period from 1998 to 2013 (see Table S1 in the supplemental material). Specifically, 398 samples were of clinical origin (sputum samples, bronchoalveolar aspirates, or *L. pneumophila* isolates), and 690 were from environmental sources (see Table S1 in the supplemental material), mainly isolates from water. Samples were obtained during routine surveillance programs for *Legionella* control as well as outbreak investigations in 97 different localities spanning 24 regions (see Fig. S1 and Table S2 in the supplemental material). All the samples were referred to our laboratory for *Legionella* detection and genetic characterization. Cultures were stored at 4°C for transportation and until being processed in our laboratory, whereas the noncultured samples were stored and kept at −20°C until they were received and processed for DNA extraction.

DNA was extracted from pure cultures by suspending bacterial biomass in water and applying a thermal shock protocol, as described previously (17). Briefly, colonies were suspended in 200 ml of 20% Chelex 100 resin (Bio-Rad Laboratories, Richmond, CA). DNA was then extracted by three freeze-thaw cycles (4°C for 5 min and 99°C for 5 min), and cellular debris was removed by pelleting at maximum speed for 1 min. Sputum samples and similar samples were treated with an UltraClean BloodSpin DNA isolation kit according to the manufacturer's instructions to extract total DNA. Quantity and purity of the nucleic acids were measured by spectrophotometry at 260 nm in triplicates by determining the A_{260}/A_{280} ratio using a NanoDrop 1000 instrument (Thermo Scientific), and DNA was stored at −20°C until use.

Sequence-based typing. The seven loci of the SBT scheme for *L. pneumophila* (*fliC*, *pilE*, *asd*, *mip*, *mompS*, *proA*, and *neuA*) (7, 8, 18) were amplified by using standard PCR. Primers, mixture, amplification, and sequencing conditions were described previously (17). A seminested PCR was subsequently applied to the sputum samples and similar samples as described previously by Coscollá and González-Candelas (19). Consensus sequences were retrieved for the amplified loci in all the samples by using forward and reverse chromatograms with the Staden package (20). The corresponding allele for each sequence was assigned based on comparisons with the *Legionella* SBT database.

SBT clustering and phylogenetic reconstruction. Isolates with complete allelic profiles (*n* = 643) were used for clustering of the different patterns into groups of single-locus variants (SLVs) by using the goeBURST full MST (minimum-spanning tree) algorithm in PHYLOViZ v1.1 (21). Strains from CV with at least four successfully genotyped loci (*n* = 778) were used for phylogenetic inference. Sequences for the SBT loci were concatenated for each isolate by using R (22). Missing loci were replaced by gaps. Maximum likelihood (ML) phylogenetic reconstruction was performed over the concatenated alignment with RAxML v7.2.8 (23), using the GTRGAMMA model of nucleotide substitution and 1,000 bootstrap replicates. The Interactive Tree of Life (iTOL) (24) utility was used to visualize and collapse the resulting phylogenetic tree.

We have previously shown that genetic variability in the genome of *L. pneumophila* strains is derived from recombination rather than substitution processes (14). This frequent exchange or acquisition of genetic material makes the actual relationships among strains better represented by a network than by a bifurcating tree. Therefore, the concatenated SBT haplotype sequences of the CV strains were used to build a median-joining network by using Network v4.613 with an epsilon value of 10 (25).

The allelic profile of the SBT scheme was downloaded for the 1,709 different sequence types available in the European Study Group for Legionella Infections (ESGLI) database at the time of starting this work (February 2014; 6,478 entries). An alignment of the seven loci for each ST was retrieved by extracting and concatenating the different alleles with R (22). The multiple-sequence alignment of the ESGLI data set was also used for ML phylogenetic reconstruction, as described above for CV data.

Analysis of diversity and population structure. Estimates of the nucleotide divergence between *L. pneumophila* populations in the two different data sets (CV and ESGLI) were obtained by using DnaSP v5.10.01 (26). The data from the ESGLI database were divided into those from European and those from non-European countries, as they are not equally represented in the database. The Ewens-Watterson test was performed by using Arlequin v3.5.1.2 (27) for the different localities and regions in CV as well as the different years of isolation. This test was applied to test whether the haplotype frequencies significantly deviated from the expected values under neutrality. Arlequin was also used for testing the partitioning of the genetic variation within and among localities and areas in the CV data set (643 isolates from 79 localities over 23 areas) as well as different regions within countries in the ESGLI data sets (2,905 samples from 171 regions over 20 European countries and 612 samples from 86 regions over 7 non-European countries) by using analysis of molecular variance (AMOVA). Non-European countries were Australia, the United States, Canada, Japan, Russia, China, and South Africa. Travel-associated cases were not considered. Bayesian modeling (BAPS software [28]) was applied in order to cluster the different STs into 15 groups (29). The resulting tree and BAPS clusters were visualized with iTOL (24).

Two factors were used as metadata to analyze the population structure of *L. pneumophila* isolates in Comunitat Valenciana: year of isolation and geographic region. xAMOVA (Daniel Wilson, unpublished data [http://www.danielwilson.me.uk/xAMOVA.html]) was performed in order to estimate the component of variance explained by each factor (two-way AMOVA). This is an extension of the original AMOVA (30) that considers the simultaneous effect of two variables on the overall variance. Statistical significance was estimated by using 10,000 permutations. Waves of the most common sequence types along the different years were studied by using R (22).

RESULTS

***L. pneumophila* diversity in the Comunitat Valenciana region.** A total of 1,088 samples were recovered from 24 different areas across CV and used for *Legionella* detection and additional SBT analysis in the period between 1998 and 2013 (see Tables S1 and S2 in the supplemental material). At least one locus could be amplified from 888 (74.5%) samples, 220 of clinical origin and 668 from environmental sources. Of the 490 isolates with a known serogroup (3 clinical and 487 environmental), 73.8% (*n* = 362) were of serogroup 1 (3 clinical and 359 environmental), known to be responsible for most of the clinical cases throughout the world (4). Of the 1,088 isolates, 643 could be successfully assigned to a specific ST (see Tables S1 and S2 in the supplemental material), and up to 778 had at least four loci sequenced and were used for further phylogenetic analyses.

From a total of 102 different STs found in the completely characterized CV data set (*n* = 643), 54 (52.9%) were found at least twice, and the rest were singletons (Fig. 1). About 55% of the strains belonged to only 9 STs. In particular, the most frequently isolated sequence type was ST1 (*n* = 203), followed by ST578 (*n* = 65), ST181 (*n* = 49), ST42 (*n* = 28), ST23 (*n* = 28), ST1358 (*n* = 21), ST2 (*n* = 13), ST269 (*n* = 11), and ST37 (*n* = 10). These STs were found spread all over the global phylogenetic tree, showing the high diversity of the *L. pneumophila* populations in CV (Fig. 2). Twenty-two different STs were found in clinical samples, representing only a subgroup of the overall environmental variability.
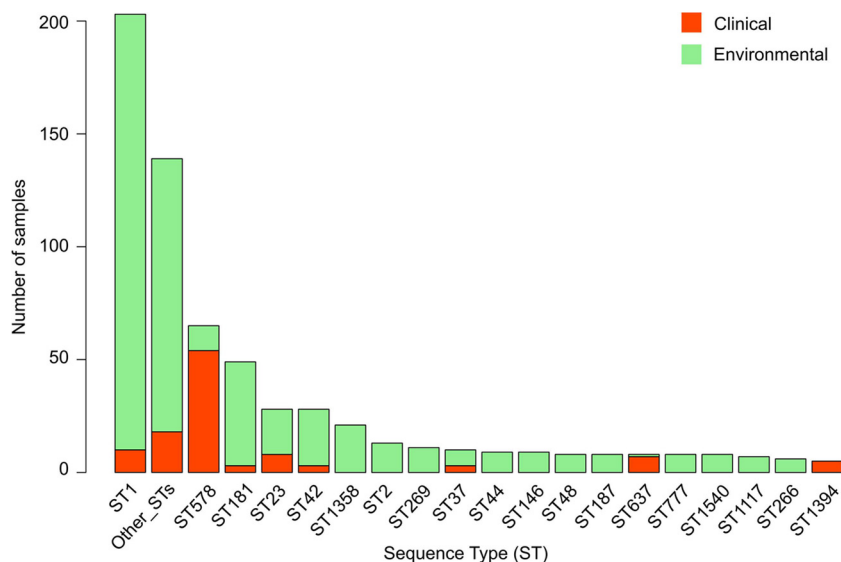
FIG 1 Bar chart representing the number of strains of clinical or environmental origin of each ST detected in CV during the period from 1998 to 2013. STs with fewer than five strains are grouped as "Other."

Twelve of these STs were also found in the environment (ST1, ST20, ST23, ST37, ST42, ST75, ST181, ST367, ST448, ST578, ST637, and ST1012). For the 10 STs found only in patients, we only had one or two isolates, except for ST1394 ($n = 5$). A median-joining network analysis (see Fig. S2 in the supplemental material) revealed the complex relationships between the different STs and suggested the existence of much more environmental variability than the one represented in our data set.

Using the SBT allelic profiles, we detected 15 complexes of single-locus variants (SLVs) derived from a specific ST that could have acted as a founder for each of the complexes (Fig. 3). The different SLVs could have been formed either by the accumulation of polymorphisms in one of the loci involved in the SBT scheme or by their involvement in recombination events.

Geographical and temporal structure. Samples with a complete ST were obtained from 23 different regions in CV. Several STs showed a wide distribution across these sampling regions (Fig. 3), with ST1 being the most widespread profile (18/23 regions), followed by ST181 (9/23 regions) and ST42 (8/23 regions). Nevertheless, ST578 was found mainly in the locality of Alcoy, where it is known to be endemic (31), and other SLVs such as ST51 or ST637 were also found exclusively in the same area. ST1 was found in 41 different localities from 18 regions, whereas 73 STs were found exclusively in different localities spread over 36 regions. Forty-eight of these STs were detected only once (singletons). The Ewens-Watterson test was performed over the 23 different regions and 79 localities and revealed that the observed haplotype frequencies in the different locations did not deviate significantly from the expected values under neutrality (P value of >0.05).

A temporal overview of the frequency of each ST per year revealed waves of STs arising in the area (Fig. 4), with ST1 being dominant in most years. The number of ST1 strains detected in the 16-year period analyzed correlated with the diversity of haplotypes found per year (Pearson's $r = 0.8644$; $P = 1.57E-05$) (see Fig. S3 in the supplemental material). Nevertheless, the years 2006 and 2009 were found to deviate from this linear relationship, mainly because of the outburst of ST181 in 2006, affecting four

different localities, and ST578 and ST42 in 2009, affecting one locality each (see Fig. S3 in the supplemental material). The ST578 peak in 2009 corresponded to an outbreak in the locality of Alcoy in 2009 (9). No further conclusion can be made from these results because of the implicit bias in the data set under study, which is dependent on the different sampling efforts made at specific points due to outbreaks or sporadic cases.

To study the contribution of year and geographical region of isolation to the distribution of the genetic variability of *L. pneumophila* in CV, we performed a two-way AMOVA (Table 1). The results showed that 33.2% of the total variance could be explained by the geographic distribution (P value of <0.0001) and that 24.6% of the variance could be explained by differences in the population structure by year (P value of <0.0001).

Analysis of local versus global variability. To better evaluate the levels of genetic diversity of *L. pneumophila* in CV, we compared the results from our data set with the global variability reported in the ESGLI database. Sequences from the seven loci of one representative of each of the 1,709 STs included in the ESGLI database were concatenated, and a ML tree was inferred (see Fig. S4 in the supplemental material). Concatenated sequences were clustered into 15 groups by using BAPS, and the results were incorporated into the phylogenetic tree. Approximately 20% of the profiles ($n = 339$) could not be assigned to specific clusters and were considered to result from admixture. Cluster 8 was the group including more profiles ($n = 413$; 24% of the total number of STs in the ESGLI data set).

Isolates from CV were present in 12 of the 15 BAPS clusters, although clinical strains spanned only 6 of them (see Fig. S4 and Table S3 in the supplemental material). Cluster 8 was also the most frequent cluster in CV ($n = 30$; 29%), containing two of the most often detected STs (ST2 and ST578), followed by clusters 11 and 12, with 24% and 13% of the CV STs, respectively (see Fig. S4 in the supplemental material). The other six most frequently reported STs clustered into five different groups, and ST269 was found as a mixed profile (see Table S3 in the supplemental material). Interestingly, one of the largest clusters in the ESGLI data set
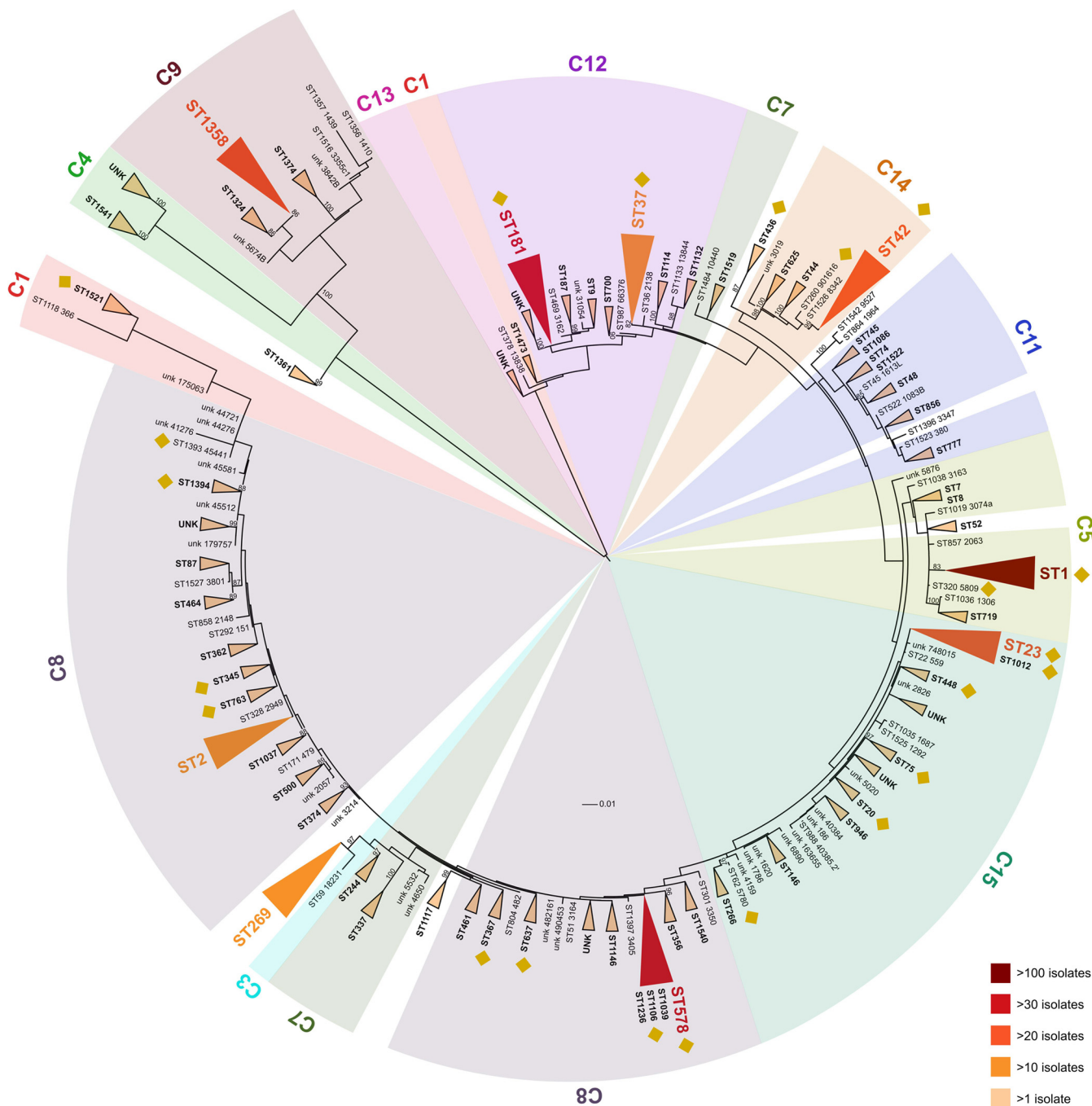
FIG 2 Maximum likelihood phylogenetic tree of the 778 *L. pneumophila* strains from CV with at least 4 SBT loci sequenced. Clades with more than one strain of the same ST have been collapsed (see the key for color differentiation of collapsed clades according to the number of samples of each ST). Shaded clades (C1 and C2, etc.) represent the different BAPS clusters, as estimated from the ESGLI data set. Yellow diamonds mark STs that have been detected in clinical cases of Legionnaires' disease. Bootstrap support values higher than 80% are shown.

(cluster 3; 110 different STs) is scarcely represented in Comunitat Valenciana, with only one ST being detected in the entire region (see Table S3 in the supplemental material). In general, the number of STs by cluster in the ESGLI data set correlated significantly with the number of STs by cluster in the local CV data set (Pearson's $r = 0.80$; P value of <0.001).

An AMOVA using the nucleotide sequences from the SBT scheme was performed with the three data sets (CV as well as

European and non-European countries from the ESGLI data sets). In the ESGLI data sets, different regions were grouped by country, while in the CV data set, different localities were grouped by area. In both cases, the highest proportion of variation was found within populations rather than among populations within groups or among groups (see Table S4 in the supplemental material). However, the distribution of variation differed between the two data sets: the ESGLI data set showed a lower proportion of varia-
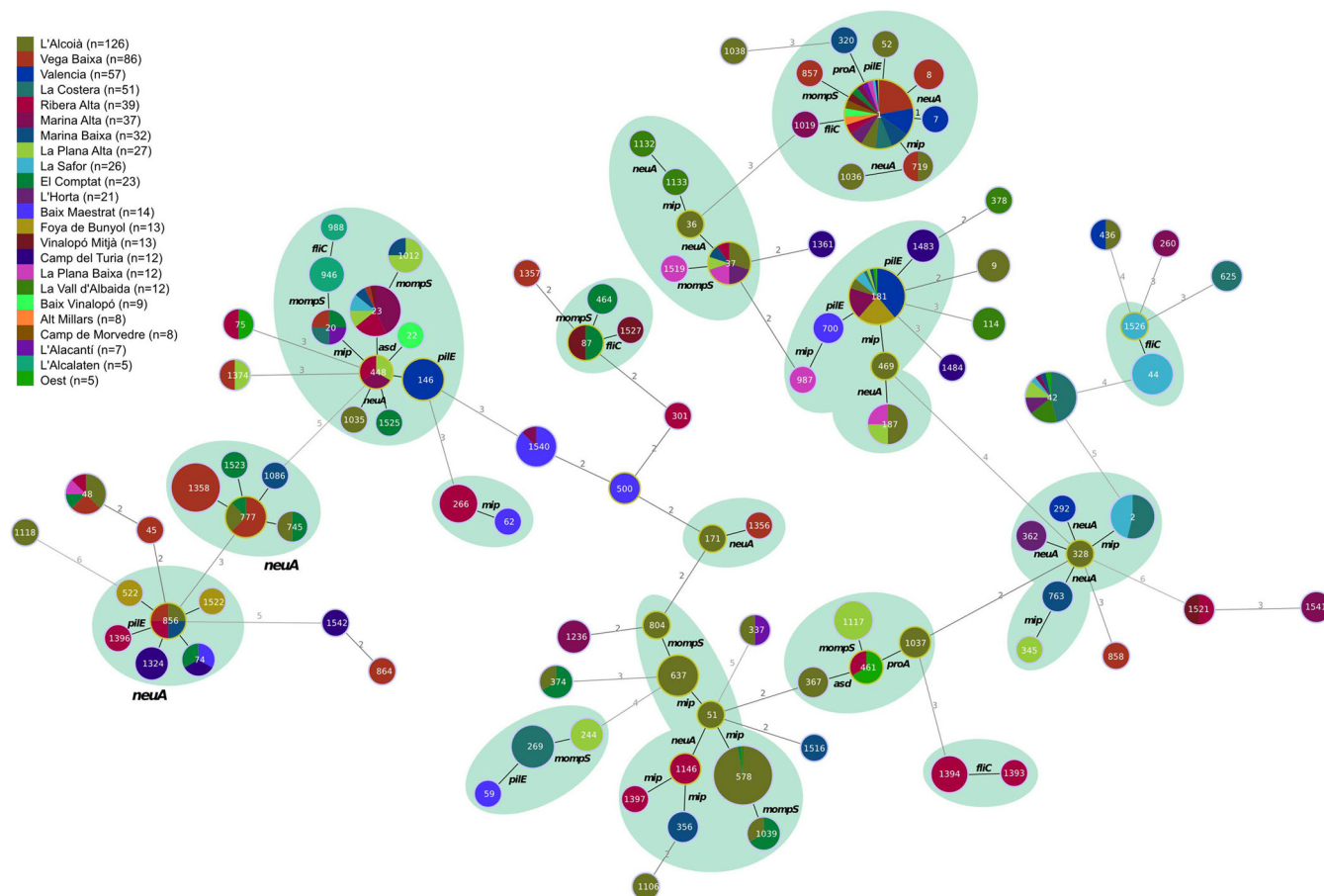
FIG 3 Minimum-spanning tree obtained with PHYLOViZ for the SBT profiles of 643 *L. pneumophila* strains. Clusters of SLVs are shaded. Pie charts represent the 102 different STs in CV, with the size being proportional to haplotype frequencies, and the different colors refer to the 23 different areas under study in the CV region (see the key). Numbers on branches represent the number of loci different from that for the founder ST. For SLVs, the name of the changing locus is represented.

tion among regions within countries (8.52% and 8.69% of variance for European and non-European countries, respectively) than among localities within areas in the CV data set (43.79% of variance). In addition, a higher proportion of the total variation was found within populations in the global data set (83.20% and 88.27% for European and non-European countries, respectively) than in CV (58.06%), reflecting the higher diversity within the first data set (see Table S4 in the supplemental material). In both cases, the highest proportion of the total variation was found within the smaller areas considered (regions within countries and localities within areas), meaning that the diversity of *L. pneumophila* found in CV is comparable to the diversity in other locations.

Genetic differentiation was assessed through the estimation of DNA divergence corrected by using the Jukes-Cantor model of nucleotide substitution. The average number of nucleotide substitutions per site between both the ESGLI and CV data sets was estimated as a Dxy(JC) value of 0.0206 ± 0.0005 (European countries) or a Dxy(JC) value of 0.0235 ± 0.0007 (non-European countries), and the number of net nucleotide substitutions per site between populations was estimated as a Da(JC) value of 0.00028 ± 0.0003 (European countries) or a Da(JC) value of 0.0006 ± 0.0007 (non-European countries). The intrapopulation nucleotide diversity estimates were comparable for all data sets

[Pi(JC, ESGLI-European) = 0.0205 ± 0.0004; Pi(JC, ESGLI-non-European) = 0.0254 ± 0.0007; Pi(JC, CV) = 0.0204 ± 0.0009]. The average numbers of nucleotide differences, *k*, between the ESGLI and CV data sets were estimated to be 50.52 and 57.25 (European and non-European countries, respectively), also similar to the intrapopulation estimates [*k*(ESGLI-European) = 49.94; *k*(ESGLI-non-European) = 61.81; *k*(CV) = 49.69]. These results further confirm that the high variability found in Comunitat Valenciana is comparable to the overall variability found in other areas worldwide.

DISCUSSION

The genetic variability of *L. pneumophila* has been assessed for countries such as the United Kingdom, Belgium, Portugal, and Italy (32–35), but there is no similar information available for Spain, despite it being significantly affected by this opportunistic pathogen. Spain is known to be one of the countries with the highest number of reported cases of Legionnaires' disease (36), and Comunitat Valenciana is the region with the second most reported cases in the country.

In a previous work, we described the variability of *L. pneumophila* in the Alicante province (south of CV) during 3 years, although only three of the current seven loci included in the SBT
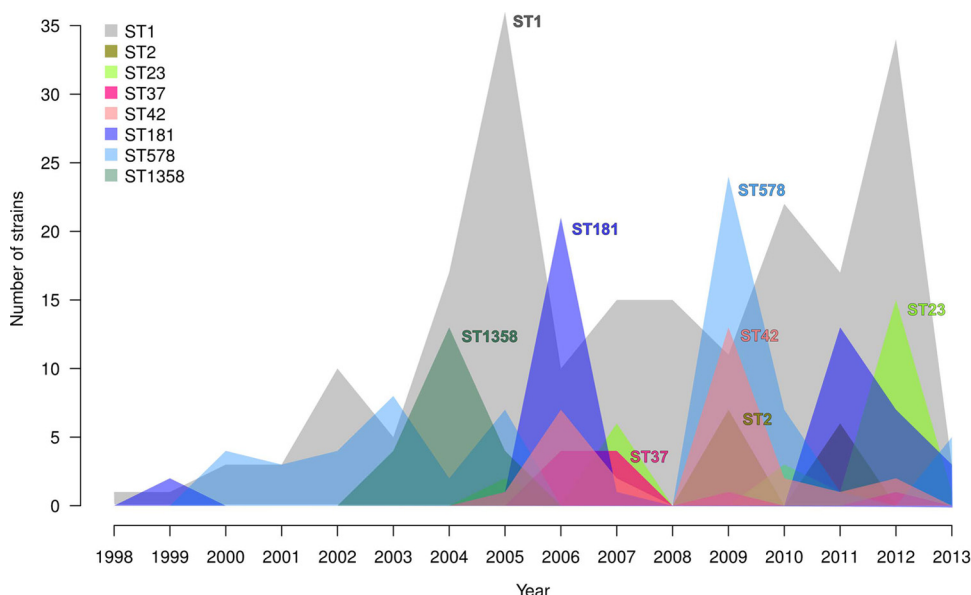
FIG 4 *L. pneumophila* STs detected in the period from 1998 to 2013 in CV, represented as numbers of strains per year. Only the eight most abundant STs that appear for more than 1 year are shown.

scheme were available at that time (12). Here, we have analyzed all the clinical and environmental strains that have been typed in our laboratory between 1998 and 2013. It is important to take into account that as the samples were obtained mainly during investigations of outbreaks and sporadic cases, there might be some bias for specific locations where these cases have occurred and where control measures have been applied in a more intense manner. However, yearly data from routine environmental surveillance programs applied over risk installations in the whole area are also included, thus balancing the possible sampling bias.

L. pneumophila is a strictly environmental pathogen, and human infection is considered accidental (15). As such, controlling this bacterium in the environment is the key point for controlling associated human infections. CV has a high rate of reported cases of legionellosis, and little is known about the population structure of this bacterium in the environment. Our results show that of the 643 samples fully typed in CV during 16 years, there is a high diversity of STs, with 102 different profiles being detected in this region. ST1 was the most frequently found ST in environmental samples, with 203/643 (31.6%) strains having this genetic profile, in agreement with data from previous reports from other areas (35, 37, 38). ST578 was found to be the second most frequently reported type (65/643; 10.1%), due mainly to the recurrent outbreaks that this ST has caused over the years in the locality of Alcoy (14). Only a subset of all CV STs was found in clinical cases, a result that has also been reported in other studies. For example, in

a surveillance work that used 443 clinical and environmental isolates from England and Wales (38), 82 different STs were detected among environmental strains, and only 42 different STs were detected in clinical isolates.

A median-joining network created by using the SBT sequence data on the CV isolates showed a complex relationship between different STs. This complexity depicts the high rate of genetic exchange among *L. pneumophila* strains (13, 14). Many nodes that represent undetected genetic profiles and whose existence is postulated in order to explain the observed relationships between the STs in the network are predicted. These unobserved profiles could be STs that have rarely caused a clinical case, are present in a very low abundance in the environment, or are uncultivable. In fact, environmental screening of water and biofilm samples using direct molecular methods has revealed the presence of mixed genetic profiles in samples resulting from the high underlying diversity present in the area (31). Besides, looking at the combination of alleles of the SBT loci, we can identify 15 complexes among the 102 STs, most of them formed by an ST that acts as a founder and SLVs that could result from intergenic recombination events from the former one.

Other surveillance works have reported that, apart from the widely distributed ST1, particular STs might be more common in specific locations, such as ST47 in the United Kingdom and Belgium (33, 38) or ST23 in Italy (35). However, these STs can also be found sporadically in other places, showing the important spread-

TABLE 1 Two-way AMOVA[a]

Factor	df	Seq SS	Adj SS	MS	*F*	Variance	% variance	*P* value
Region	22	3,835.2	5,434.1	247.00	17.615	10.99	33.15	<0.001
Yr	16	3,644.5	3,644.5	227.78	16.244	8.15	24.58	<0.001
Others	604	8,469.6	8,469.6	14.02		14.02	42.27	

[a] The effect of geographical and temporal structuring on the genetic variability of *L. pneumophila* samples from Comunitat Valenciana in the 1998-2013 period was evaluated. df, degree of freedom; Seq SS, sequential sum of square deviations for each factor; Adj SS, adjusted sum of square deviations for each factor; MS, mean square deviation (Adj SS/df); *F*, *F* statistic representing the ratio between each factor (region/year) and other factors.

ing potential of this environmental pathogen. For instance, ST23 was found to cause an important outbreak in a hotel in the coastal city of Calpe (Alicante, Spain) in 2012 (39). Our analyses show that ~50% of the observed variability can result from geographical (33.2%) or temporal (24.6%) structuring (Table 1). Although 29 STs were found in more than one locality, the majority of profiles (73 STs, considering the 48 singletons) were found exclusively in specific localities spanning 23 different regions of CV. However, no significant deviation from the expected haplotype diversity was found in the different localities and regions. The same result was obtained for samples taken in different years. Temporal structuring was found in the shape of waves of STs bursting in specific years, such as ST181 in 2006 or ST578 in 2009. Nevertheless, this result can be affected by the above-mentioned sampling bias, because the high numbers of these STs are dependent on the sampling effort in these particular years due to outbreaks. ST181 was found in four different localities in 2006, so we can say more confidently that there was an outburst of this ST in the area. All the ST578 cases in 2009 corresponded to a specific outbreak that occurred at that time in the locality of Alcoy (9).

The comparison between the local (CV) and global (ESGLI) diversity of *L. pneumophila* revealed large similarities and little genetic differentiation between the two data sets. For instance, we mapped the different STs found in CV onto the global tree (see Fig. S4 in the supplemental material) and found environmental representatives from CV for 12 out of 15 clusters. STs found in clinical cases were included in only 6 of these clusters. In general, a good correlation was found between the numbers of STs from each data set in each cluster. An analysis of variance over the global data set, encompassing 171 regions in 20 European countries and 86 regions in 7 non-European countries, showed that most of the variability was found within the different populations, meaning that there is a high diversity of *L. pneumophila* in most of the locations included in the test. A comparable result was obtained for 79 localities spanning 23 areas in CV. These results support a previous observation that ~80% of the total genetic variability in a specific area within CV could be attributed to intrapopulation differences (12). The level of nucleotide differentiation between the two data sets is similar to the level of genetic diversity within each data set. These results indicate that two strains from the same data set are almost as likely to be as different as any two other strains from different data sets each. Furthermore, these high levels of variability are found mostly at the smallest geographic scales tested in each data set and in many different places. Interestingly, this variability is due mainly to variants isolated only once, which might imply that even a higher diversity from undetected or uncultivable environmental strains remains yet undiscovered.

Probably, the main weakness of this work derives from the sampling design. We have combined clinical and environmental samples in the same analysis, which can be justified by the dead-end nature of *Legionella* infection in humans. In consequence, a clinical isolate was also an environmental sample a short time before it infected a person, and in this respect, both clinical and environmental samples are representative of the total *Legionella* population in an area. However, the sampling efforts are very different in areas with high incidences of legionellosis outbreaks and community cases, such as Comunitat Valenciana, because most samplings are driven by investigations of these cases. In this particular study, we have collaborated with public health authorities to include environmental samples from routine, non-out-break-related surveillance analyses, thus widening the opportunities for analyzing the environmental variability of *L. pneumophila* in this area. However, similar efforts are not common, and the information being deposited in public databases such as the ESGLI database is still dependent on analyses of clinical isolates and the occasional environmental samples retrieved during investigations of corresponding outbreaks or community cases.

In conclusion, we have characterized the genetic diversity of clinical and environmental samples of *Legionella pneumophila* from the Comunitat Valenciana region in Spain from 1998 to 2013, and we have compared it with the overall variability reported in the ESGLI database (including mainly samples from European countries). Although some STs have been found in specific localities, the most abundant STs can be found repeatedly in different countries. The high genetic variability detected can be created through the exchange of genetic material between strains that can spread to other locations. *L. pneumophila* is thought to be a recent pathogen for humans, dispersing through aerosols and for which no person-to-person transmission has been reported, facts that contrast with the wide distribution of this pathogen. Genomic and metagenomic analyses will be crucial to evaluate the fitness of different lineages under distinct environmental conditions, and more intense sampling projects are likely to find additional reservoirs that explain the as-yet-undetected genetic diversity of this bacterium. Furthermore, this information will be crucial to elucidate the main factors responsible for the observed global distribution of some STs and the adaptive or accidental nature of the minority variants found at local scales.

ACKNOWLEDGMENTS

REFERENCES

1. **Donlan RM.** 2002. Biofilms: microbial life on surfaces. Emerg Infect Dis **8:**881–890. http://dx.doi.org/10.3201/eid0809.020063.
2. **Declerck P.** 2010. Biofilms: the environmental playground of *Legionella pneumophila*. Environ Microbiol **12:**557–566. http://dx.doi.org/10.1111/j.1462-2920.2009.02025.x.
3. **Declerck P, Behets J, Margineanu A, van Hoef V, De Keersmaecker B, Ollevier F.** 2009. Replication of *Legionella pneumophila* in biofilms of water distribution pipes. Microbiol Res **164:**593–603. http://dx.doi.org/10.1016/j.micres.2007.06.001.
4. **Fields BS, Benson RF, Besser RE.** 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. Clin Microbiol Rev **15:**506–526. http://dx.doi.org/10.1128/CMR.15.3.506-526.2002.
5. **Fraser DW, Tsai TR, Orenstein W, Parkin WE, Beecham HJ, Sharrar RG, Harris J, Mallison GF, Martin SM, McDade JE, Shepard CC, Brachman PS.** 1977. Legionnaires' disease: description of an epidemic of pneumonia. N Engl J Med **297:**1189–1197. http://dx.doi.org/10.1056/NEJM197712012972201.
6. **Lück C, Fry NK, Helbig JH, Jarraud S, Harrison TG.** 2013. Typing methods for *Legionella*. Methods Mol Biol **954:**119–148. http://dx.doi.org/10.1007/978-1-62703-161-5_6.
7. **Gaia V, Fry NK, Afshar B, Lück PC, Meugnier H, Etienne J, Peduzzi R, Harrison TG.** 2005. Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumo-*

phila. J Clin Microbiol **43**:2047–2052. http://dx.doi.org/10.1128/JCM.43
.5.2047-2052.2005.

8. **Ratzow S, Gaia V, Helbig JH, Fry NK, Lück CP, Helbig HJ.** 2007.
Addition of *neuA*, the gene encoding *N*-acylneuraminate cytidylyl trans-
ferase, increases the discriminatory ability of the consensus sequence-
based scheme for typing *Legionella pneumophila* serogroup 1 strains. J Clin
Microbiol **45**:1965–1968. http://dx.doi.org/10.1128/JCM.00261-07.

9. **Coscollá M, Fenollar J, Escribano I, González-Candelas F.** 2010. Legionel-
losis outbreak associated with asphalt paving machine, Spain, 2009. Emerg
Infect Dis **16**:1381–1387. http://dx.doi.org/10.3201/eid1609.100248.

10. **White PS, Graham FF, Harte DJG, Baker MG, Ambrose CD, Humphrey
RG.** 2013. Epidemiological investigation of a Legionnaires' disease out-
break in Christchurch, New Zealand: the value of spatial methods for
practical public health. Epidemiol Infect **141**:789–799. http://dx.doi.org
/10.1017/S0950268812000994.

11. **Keramarou M, Evans MR, South Wales Legionnaires' Disease Outbreak
Control Team.** 2010. A community outbreak of Legionnaires' disease in
South Wales, August-September 2010. Euro Surveill **15**(42):pii=19691.
http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19691.

12. **Coscollá M, Gosalbes MJ, Catalán V, González-Candelas F.** 2006. Ge-
netic variability in environmental isolates of *Legionella pneumophila* from
Comunidad Valenciana (Spain). Environ Microbiol **8**:1056–1063. http:
//dx.doi.org/10.1111/j.1462-2920.2006.00997.x.

13. **Coscollá M, González-Candelas F.** 2007. Population structure and recom-
bination in environmental isolates of *Legionella pneumophila*. Environ Mi-
crobiol **9**:643–656. http://dx.doi.org/10.1111/j.1462-2920.2006.01184.x.

14. **Sánchez-Busó L, Comas I, Jorques G, González-Candelas F.** 2014.
Recombination drives genome evolution in outbreak-related *Legionella
pneumophila* isolates. Nat Genet **46**:1205–1211. http://dx.doi.org/10.1038
/ng.3114.

15. **Mekkour M, Driss EKB, Tai J, Cohen N.** 2013. *Legionella pneumophila*:
an environmental organism and accidental pathogen. Int J Sci Technol
2:187–196.

16. **Instituto de Salud Carlos III.** 2012. Brotes de legionelosis notificados
a la Red Nacional de Vigilancia Epidemiológica en el periodo 1999–
2011. Instituto de Salud Carlos III, Madrid, Spain. http://www.isciii.es
/ISCIII/es/contenidos/fd-servicios-cientifico-tecnicos/fd-vigilancias-
alertas/fd-enfermedades/legionelosis.shtml.

17. **Sánchez-Busó L, Coscollá M, Pinto-Carbó M, Catalán V, González-
Candelas F.** 2013. Genetic characterization of *Legionella pneumophila*
isolated from a common watershed in Comunidad Valenciana, Spain.
PLoS One **8**:e61564. http://dx.doi.org/10.1371/journal.pone.0061564.

18. **Gaia V, Fry NK, Harrison TG, Peduzzi R.** 2003. Sequence-based typing
of *Legionella pneumophila* serogroup 1 offers the potential for true porta-
bility in legionellosis outbreak investigation. J Clin Microbiol **41**:2932–
2939. http://dx.doi.org/10.1128/JCM.41.7.2932-2939.2003.

19. **Coscollá M, González-Candelas F.** 2009. Direct sequencing of *Legionella
pneumophila* from respiratory samples for sequence-based typing analysis.
J Clin Microbiol **47**:2901–2905. http://dx.doi.org/10.1128/JCM.00268-09.

20. **Staden R.** 1996. The Staden sequence analysis package. Mol Biotechnol
5:233–241. http://dx.doi.org/10.1007/BF02900361.

21. **Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M,
Carriço JA.** 2012. PHYLOViZ: phylogenetic inference and data visualiza-
tion for sequence based typing methods. BMC Bioinformatics **13**:87. http:
//dx.doi.org/10.1186/1471-2105-13-87.

22. **R. Core Team.** 2014. R: a language and environment for statistical com-
puting. R Foundation for Statistical Computing, Vienna, Austria. http:
//www.r-project.org/.

23. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phy-
logenetic analyses with thousands of taxa and mixed models. Bioinformat-
ics **22**:2688–2690. http://dx.doi.org/10.1093/bioinformatics/btl446.

24. **Letunic I, Bork P.** 2011. Interactive Tree of Life v2: online annotation and

25. **Bandelt HJ, Forster P, Röhl A.** 1999. Median-joining networks for in-
ferring intraspecific phylogenies. Mol Biol Evol **16**:37–48. http://dx.doi
.org/10.1093/oxfordjournals.molbev.a026036.

26. **Librado P, Rozas J.** 2009. DnaSP v5: a software for comprehensive anal-
ysis of DNA polymorphism data. Bioinformatics **25**:1451–1452. http://dx
.doi.org/10.1093/bioinformatics/btp187.

27. **Excoffier L, Lischer H.** 2010. Arlequin suite ver 3.5: a new series of
programs to perform population genetics analyses under Linux and
Windows. Mol Ecol Resour **10**:564–567. http://dx.doi.org/10.1111/j
.1755-0998.2010.02847.x.

28. **Corander J, Marttinen P, Sirén J, Tang J.** 2008. Enhanced Bayesian
modelling in BAPS software for learning genetic structures of populations.
BMC Bioinformatics **9**:539. http://dx.doi.org/10.1186/1471-2105-9-539.

29. **Underwood AP, Jones G, Mentasti M, Fry NK, Harrison TG.** 2013.
Comparison of the *Legionella pneumophila* population structure as deter-
mined by sequence-based typing and whole genome sequencing. BMC
Microbiol **13**:302. http://dx.doi.org/10.1186/1471-2180-13-302.

30. **Excoffier L, Smouse PE, Quattro JM.** 1992. Analysis of molecular vari-
ance inferred from metric distances among DNA haplotypes: application
to human mitochondrial DNA restriction data. Genetics **131**:479–491.

31. **Sánchez-Busó L, Olmos MP, Camaró ML, Adrián F, Calafat JM,
González-Candelas F.** 2015. Phylogenetic analysis of environmental *Le-
gionella pneumophila* isolates from an endemic area (Alcoy, Spain). Infect
Genet Evol **30**:45–54. http://dx.doi.org/10.1016/j.meegid.2014.12.008.

32. **Harrison TG, Doshi N, Fry NK, Joseph C.** 2007. Comparison of clinical
and environmental isolates of *Legionella pneumophila* obtained in the UK
over 19 years. Clin Microbiol Infect **13**:78–85. http://dx.doi.org/10.1111
/j.1469-0691.2006.01558.x.

33. **Vekens E, Soetens O, De Mendonça R, Echahidi F, Roisin S, Deplano A,
Eeckhout L, Achtergael W, Piérard D, Denis O, Wybo I.** 2012. Sequence-
based typing of *Legionella pneumophila* serogroup 1 clinical isolates from
Belgium between 2000 and 2010. Euro Surveill **17**(43):pii=20302. http:
//www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20302.

34. **Chasqueira MJ, Rodrigues L, Nascimento M, Marques T.** 2009. Sequence-
based and monoclonal antibody typing of *Legionella pneumophila* isolated
from patients in Portugal during 1987-2008. Euro Surveill **14**(28):pii=19271.
http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19271.

35. **Fontana S, Scaturro M, Rota MC, Caporali MG, Ricci ML.** 2014.
Molecular typing of *Legionella pneumophila* serogroup 1 clinical strains
isolated in Italy. Int J Med Microbiol **304**:597–602. http://dx.doi.org/10
.1016/j.ijmm.2014.04.004.

36. **Beauté J, Zucs P, de Jong B, European Legionnaires' Disease Surveil-
lance Network.** 2013. Legionnaires' disease in Europe, 2009-2010. Euro
Surveill **18**(10):pii=20417. http://www.eurosurveillance.org/ViewArticle
.aspx?ArticleId=20417.

37. **Amemura-Maekawa J, Kura F, Chang B, Watanabe H.** 2005. *Legionella
pneumophila* serogroup 1 isolates from cooling towers in Japan form a
distinct genetic cluster. Microbiol Immunol **49**:1027–1033. http://dx.doi
.org/10.1111/j.1348-0421.2005.tb03699.x.

38. **Harrison TG, Afshar B, Doshi N, Fry NK, Lee JV.** 2009. Distribution of
Legionella pneumophila serogroups, monoclonal antibody subgroups and
DNA sequence types in recent clinical and environmental isolates from
England and Wales (2000–2008). Eur J Clin Microbiol Infect Dis **28**:781–
791. http://dx.doi.org/10.1007/s10096-009-0705-9.

39. **Vanaclocha H, Guiral S, Morera V, Calatayud MA, Castellanos M,
Moya V, Jerez G, Gonzalez F.** 2012. Preliminary report: outbreak of
Legionnaires disease in a hotel in Calp, Spain, update on 22 February 2012.
Euro Surveill **17**(8):pii=20093. http://www.eurosurveillance.org/View
Article.aspx?ArticleId=20093.